(12) UK Patent Application (19) GB (11) 2 342 527 (13) A

(43) Date of A Publication 12.04.2000

(21) Application No 9821554.4

(22) Date of Filing 02.10.1998

(71) Applicant(s)
General Datacomm Inc
(Incorporated in USA - Delaware)
Straits Turnpike, Route 63, Middlebury,
Connecticut 06762-1299, United States of America

(72) Inventor(s)
David Banes

(74) Agent and/or Address for Service
Mathys & Squire
100 Grays Inn Road, LONDON, WC1X 8AL,
United Kingdom

(51) INT CL[7]
H04L 12/56

(52) UK CL (Edition R)
H4K KTK
H4P PPS

(56) Documents Cited
GB 2261799 A EP 0544454 A EP 0540028 A
US 5485457 A US 5337308 A

(58) Field of Search
UK CL (Edition R) H4K KTK , H4P PPS
INT CL[7] H04L 12/56
ONLINE : WPI ; EPODOC ; JAPIO

(54) Abstract Title
Data switch performance monitoring

(57) In a data switch such as an ATM switch, delay within a switching core is measured by stamping a data packet with time information on an input side of the switch core and reading the time stamp information on the output side of the switch core. The time stamp is preferably inserted in a dummy header and may be applied to cells of a particular characteristic, for example on a particular connection.

GB 2 342 527 A

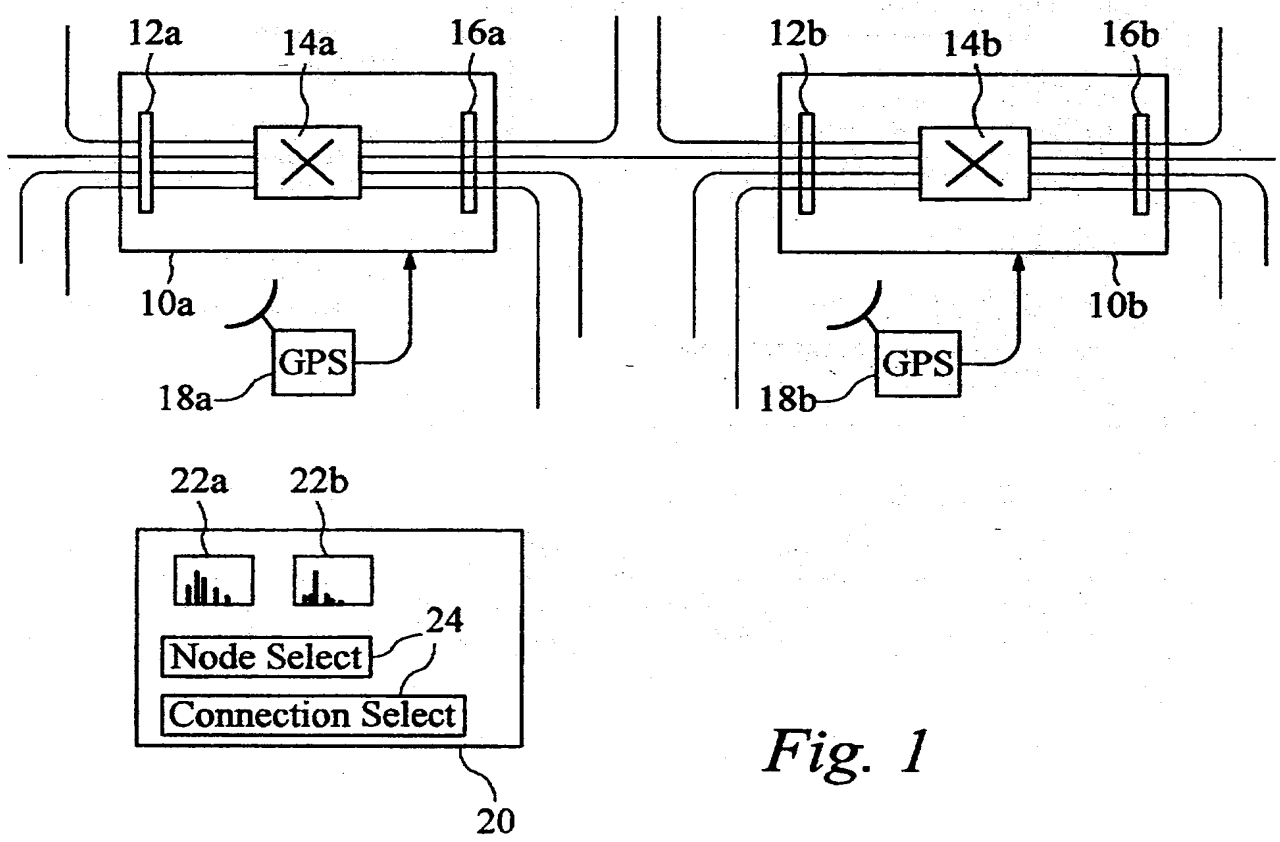

Fig. 1

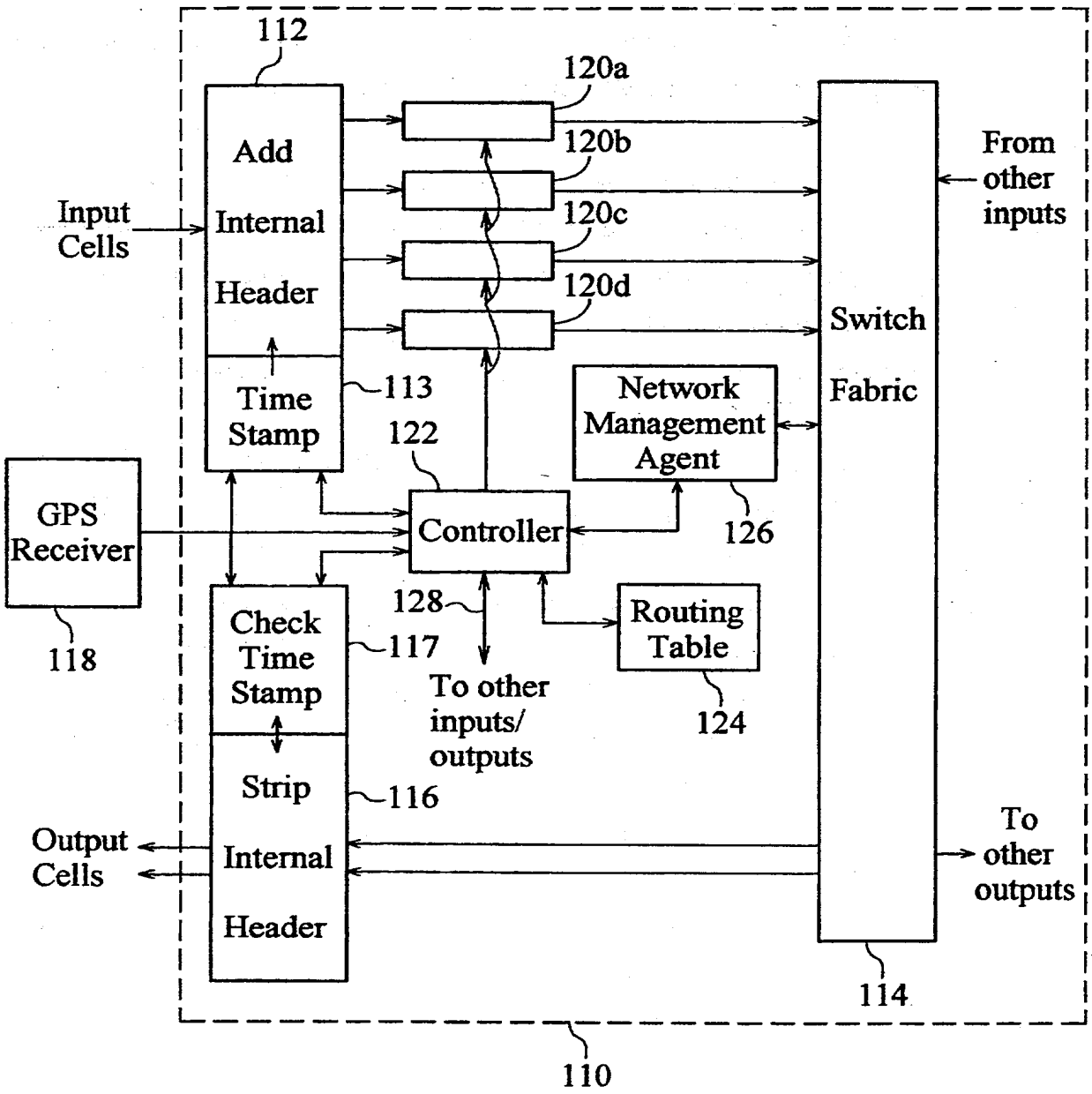

Fig. 2

DATA SWITCH PERFORMANCE MONITORING

The present invention relates to data switches, particularly, but not exclusively, packet switches such as ATM switches.

In a communication network comprising a series of data switches or nodes, it is often desirable to monitor network performance, for example by monitoring parameters such as overall delay across a communication link and the variations in delay. This may be necessary in order to ensure that agreed performance criteria for a particular connection are met.

Conventionally, this is achieved by sending additional data, for example special cells or packets (the terms "cell" and "packet" are used interchangeably herein to denote discrete quantities of data together with any associated addressing or control information) and monitoring their transit through the network. For example a monitoring cell may be sent across the network carrying data encoding a transmission time and the time of receipt at the destination may be recorded to give a measure of average transit delay. The measurement may be repeated for a number of cells, to give measurement of variations in transit delay.

A problem that has been appreciated with this method is that the insertion of such cells uses network bandwidth and also may give unreliable measurements, as the very act of measuring the transit delay may alter the measurement. Another problem the inventor has appreciated is that if specially identified cells are used to carry the information, for example Operations And Maintenance (OAM) cells in an ATM network, these may be handled differently from "normal" data cells within switches within the network, and the transit delay for these cells may be markedly different from the transit delay for normal data traffic. Another problem the inventor has identified is that with such methods, if congestion is identified, it is not easy to determine exactly where the congestion is occurring.

The present invention aims to alleviate the above problems.

In a first aspect, the invention provides a data switch including a switch core for switching packets of data, the switch comprising, on an input side of the switch core, means for adding time stamp information to one or more data packets to be switched; and, on an output side of the switch core means for obtaining a measure of the delay of the or each packet within the switch based on the time stamp information.

In this way, a measure of the actual transit delay for genuine data cells (as opposed to inserted monitoring cells) can be obtained for a specific connection in a specific switch in the network. This may allow the source of any congestion to be pinpointed much more readily.

Preferably, the switch is arranged to add time stamp information, or at least a dummy header corresponding to the time stamp information to all data cells, or at least all data cells of specified characteristics, passing through the switch core. In this way, the act of time-stamping a particular cell should not alter the measurement, so the readings may be more reliable. In addition, if all cells are time stamped similarly, processing of the cells may be more uniform, and hence the switch architecture may be simplified.

Preferably, the switch has means for selectively monitoring transit times of cells having defined characteristics. For example, the monitoring may be configurable to monitor all cells defining a particular route through the switch, or all cells originating from a given input or a given output. Where the data communication protocol is of a type which supports virtual connections, for example an ATM network in which virtual paths (VPs) and virtual channels (VCs) are defined, the monitoring means is preferably configurable to monitor transit times for cells (either all cells or a defined proportion of cells) having defined VPIs and/or VCIs.

Preferably, the switch has means for performing a plurality of independent monitoring operations concurrently. For example, the switch may include means for selecting a first traffic stream to monitor corresponding to cells conforming to a first set of defined criteria (for example a defined VCI and/or VPI) and a second traffic stream to monitor corresponding to cells conforming to a second set of defined criteria. The data switch will normally include routing table means for storing information controlling the routing of cells within the switch. Preferably, the routing table means is arranged to store information controlling the monitoring of cells. For example, in an ATM switch, there will normally be a routing table containing information controlling translation of VCI and VPI values and physical routing of cells from inputs to outputs, as well as policing parameters. It is preferable if the routing table is arranged to store additional information (for example one or more flag bits) specifying whether or not cells on a particular connection are to be monitored, and, in the case of more than one independent monitoring operation to be performed, an identifier of the monitoring operation associated with the cells. This allows an operator great flexibility in choosing which cells to monitor, and can simplify identification of problems. For example, at a simple level, cells on a single VC could be monitored in a single monitoring operation. In addition cells on multiple VCs, having a common characteristic, for example all originating from a common source, or travelling, at least at some stage in the network, along a common link, may be monitored in a separate monitoring operation.

Preferably, the monitoring means is arranged to increment at least one counter value in dependence on the measure of transit delay for the cell. In a simple implementation, the monitoring means may be arranged to increment a single counter value for each cell for which the transit delay exceeds a predetermined threshold.

More preferably, however, the monitoring means is arranged to increment one of a plurality of counters, each counter corresponding to a

predetermined range of transit delays. This may allow a histogram of cell transit times to be monitored, from which a measure of variation in cell transit times can be determined, in addition to determining a proportion of cells which exceed a desired transit time.

In a preferred implementation, the data switch is associated with means for displaying transit time values graphically.

The data switch may be associated with a network controller for controlling a plurality of data switches, the network controller being arranged to configure each data switch to monitor cells of a defined characteristic and to obtain the results of monitoring from each data switch. Preferably, the network controller includes means for inputting an identifier of a connection across the network to be monitored and is arranged to configure each data switch associated with that connection to monitor cells associated with the connection and to return the results of monitoring to the network controller. The network controller is most preferably arranged to provide a visual output of a representation of each network node and the transit delay data associated with that network node (preferably as a histogram associated with each node, but optionally numerically or by means of other visual symbols).

Preferably, in addition to obtaining a measure of transit delay within the data switch (intra-node delay), the network controller is arranged to determine a measure of transit delay between network nodes (inter-node delay). This may be achieved either in the conventional manner by sending one or more special cells across the network, or by correlating the time stamps assigned to cells in each switch. Preferably, correlation of the time stamps comprises communicating the time stamp and an identifier of the cell to which it relates to a controller, preferably the network controller. The local time used to generate the time stamp at each data switch is preferably synchronised, or a table of absolute off-set values relative to a reference time may be stored.

GPS time, or other universal time reference, may be input and employed to correlate cell transit times.

Communication between switches and a network controller may be effected by dedicated communications links, or by means of the network itself. A protocol such as SNMP (Simple Network Management Protocol) may be employed.

The transit times determined within the switch may be provided as an input to means for determining queuing order within the switch; in this way queuing can be adjusted dynamically to take into account actual delays within a switch. The priority level assigned to a queue may be adjusted in response to a measure of actual transit time for cells associated with that queue. In particular, in response to detection that transit delays are exceeding a predetermined threshold (for example an agreed delay time) for a particular queue or connection, the connection or queue may be assigned a high or maximum priority.

The invention extends to methods of operation and to networks implementing time stamping in at least some switches. Further aspects are set out in the claims.

An embodiment of the invention will now be described, by way of example, with reference to the accompanying drawings in which:

Fig. 1 is a schematic view of a portion of a simplified network according to a first embodiment; and

Fig. 2 is a schematic view of a switch according to a further embodiment.

Referring to Fig. 1, a simplified network comprises first and second data switches 10a,10b which are connected to each other and to other switches and nodes (not shown). For ease of understanding, a unidirectional network will be described (in practice, a bi-directional network can be composed of paired unidirectional connections, as is well known); it is to be assumed that data flows from left to right in the diagram.

Each switch 10a, 10b includes, on an input side, a time stamper 12a, 12b for adding a time stamp to data packets as they enter the switch. The stamped packets are passed to a switch core 14a, 14b where they are routed appropriately and emerging data packets pass through a time stamp reader 16a, 16b which determines a measure of transit delay from the difference between the time at which the stamp is read and the stored time. To give an accurate measure of cell delay within the switch, each time stamper 12a, 12b should ideally be located (logically) as close as practical to the input of the switch, and preferably before any buffering or at least before substantial buffering or buffering which introduces a significantly variable delay. Likewise, each time stamp reader 16a, 16b should be located (logically) as close as practical to the output, preferably after all substantial buffering.

Each switch is arranged to receive an input of GPS time from a respective GPS receiver 18a, 18b. The input GPS time is supplied via a controller (not shown in Fig. 1) to the time stampers 12a, 12b. The time stamp applied to a particular cell is read by readers 16a, 16b when the cell emerges from the switch fabric and compared to the current time. By comparing the times of input and egress, a measure of cell transit delay within the cell (intra-node delay) can be obtained. Furthermore, by correlating times between switches, a measure of inter-node delay can be obtained; this may be achieved by encoding the GPS time value in the cell contents for transmission across the network, or by passing information relating to cells between switches by some other means, for example by means of network

management messages.

In the embodiment schematically depicted, information collected from a number of switches is passed to a network controller 20, typically over the network itself, for example using Simple Network Management Protocol (SNMP). This is arranged to display graphically, for example, a histogram 22a shown variation of cell delays within a given switch and a histogram 22b shown variation of cell delays on a given connection. Inputs 24 are provided to enable a user to select the monitoring operation to be performed.

Referring to Fig. 2, a switch according to a further embodiment, in which weighted fair queuing is employed, will now be described.

A switch 110 is arranged to receive input ATM cells, for example from a fibre-optic medium, and add an internal header to the cells, to facilitate routing within the cell, in a pre-processor 112. The details of this may be based, for example, on arrangements described in our earlier UK patent applications nos 9810076.1, 9810080.3 and 9810102.5, the disclosures of each of which are incorporated herein by reference. In addition to adding a header containing internal routing information, the cell pre-processor 112 adds a time stamp supplied by time stamper 113. The time stamper 113 is under the control of controller 122 and can be set to stamp only cells of a defined type, for example based on information stored in routing table 124.

As an example of selective monitoring, based on information within the routing table, the time stamper may stamp only cells having particular VCI or VPI values, or may assign a particular flag value to cells of particular VCI or VPI values so that those cells may be later distinguished and monitored as a specific group.

The cells are supplied to a plurality of queues 120a...120d for buffering prior to passing to switch fabric 114. In the simplified embodiment depicted,

there are four queues shown, but there may be many more queues in a practical implementation. For example, there may be one queue per output of the switch fabric. More preferably, there is one queue per connection, or at least several queues per switch fabric output, one for each class of connection. Particularly where large and variable numbers of queues are provided, the queues are preferably allocated dynamically in software.

The order in which cells are output from the queues will depend on switch fabric availability, but is under the influence of controller 122, which is initially set to implement a predetermined priority scheme.

After routing within the switch fabric 114, the cells emerge to one of a plurality of post-processors, of which a single example 116 is depicted. The post-processor removes the internal header and passes the cells to, in this embodiment, one of two outputs, the output being selected in dependence on the contents of the internal header. The post-processor incorporated a time-stamp reader 117 which reads any time stamps on cells emerging and communicates information based on this, together with an identifier of the cell where required to the controller 122.

It will be appreciated that an incoming cell may in principle be directed by the switch fabric to any of the outputs. To monitor transit delays, it is therefore necessary to measure time stamps at each output; this may be achieved by multiple controllers or a single controller within the switch communicating with each input and output interface. In the embodiment depicted, the controller has a link 128 to other interfaces and is able to monitor all inputs and outputs. In a practical implementation, there may be a central system controller performing overall supervision of the switch, with certain monitoring tasks delegated to separate slot controllers, each associated with one input or output or a subset of the inputs and outputs.

The controller 122 is arranged to modify the priority scheme to take into

account measured cell delays. Where cells are queued on a per-connection basis and monitoring is performed for each connection, accurate control of delays on each connection may be effected within tight tolerance limits. In other cases, the priority scheme may be altered only in case of exception, for example when it is determined that cells in a particular queue (however that queue is organised) are becoming unduly delayed.

Because more direct control of transit delays is possible, novel services, in place of conventional CBR, VBR, ABR, etc services may be implemented. For example, a guaranteed throughput may be specified.

To effect control and monitoring of cell delays across a network and to use the information collected within the switch for network management purposes, it is necessary for the controller to communicate with the world outside the switch. This is effected by means of network management agent 126, for example running SNMP, which conveys the information generated by the controller to external network management devices. In the embodiment depicted, the network management agent communicates with the switch fabric 114 directly, for inserting and receiving cells, for example OAM cells, for communicating across the data network in which the switch operates. It is possible, however, for dedicated control links to be provided independently of the data links.

Where inter-node delays are to monitored, the network management agent may send information enabling particular cells to be "tracked" from node to node. This may be achieved by using special cells encoded in a particular way; however, this requires additional cells and runs the risk that the special cells will encounter different delays to "real" cells which it is desired to monitor. Another option is to calculate a "signature" for one or more cells, and communicate this signature together with cell arrival or departure time. Then, by comparing departure and arrival times for cells having matching signatures at different nodes, transit delays across the network may be

monitored. The signature should be sufficiently distinctive to minimise the chance of a false match; even so, it is desirable to match signatures for a sequence of cells, so that spurious matches can be isolated. A four octet signature (32 bits), provided it is suitably chosen (for example based on a CRC algorithm) should theoretically produce a false match every 2^32 cells (approximately 1 in 4*10^9). Given the very high cell throughput rates of a typical switch, isolated false matches will still occurr, but the chances of a sequence of false matches occuring are remote.

The examples described above enable intra-node and inter-node delays to be monitored and enable queuing to be adjusted to take into account actual delays. Other applications and implementations of the invention are possible, and the invention is not to be construed as limited to the examples described above.

Each feature described herein may be independently provided, unless otherwise stated.

Claims

1. A data switch including a switch core for switching packets of data, the switch comprising, on an input side of the switch core, means for adding time stamp information to one or more data packets to be switched; and, on an output side of the switch core means for obtaining a measure of the delay of the or each packet within the switch based on the time stamp information.

2. A switch according to Claim 1 arranged to add time stamp information, or at least a dummy header corresponding to the time stamp information, to all data cells of specified characteristics passing through the switch core.

3. A switch according to Claim 1 or 2, having means for selectively monitoring transit times of cells having defined characteristics.

4. A switch according to Claim 3, wherein the monitoring means is configurable to monitor transit times for cells having defined VPIs and/or VCIs.

5. A switch according to any preceding claim having means for performing a plurality of independent monitoring operations concurrently.

6. A switch according to any preceding claim including routing table means for storing information controlling the routing of cells within the switch and further arranged to store information controlling the monitoring of cells.

7. A switch according to Claim 6 wherein the routing table is arranged to store additional information specifying whether or not cells on a particular connection are to be monitored, and preferably an identifier of the monitoring

operation associated with the cells.

8. A switch according to any preceding claim wherein the monitoring means is arranged to increment at least one counter value in dependence on the measure of transit delay for the cell.

9. A switch according to Claim 8 wherein the monitoring means is arranged to increment one of a plurality of counters, each counter corresponding to a predetermined range of transit delays.

10. An arrangement comprising a data switch according to any preceding claim associated with means for displaying transit time values graphically.

11. An arrangement comprising a data switch according to any of Claims 1 to 9 associated with a network controller for controlling a plurality of data switches, the network controller being arranged to configure each data switch to monitor cells of a defined characteristic and to obtain the results of monitoring from each data switch.

12. An arrangement according to Claim 11, wherein the network controller includes means for inputting an identifier of a connection across the network to be monitored and is arranged to configure each data switch associated with that connection to monitor cells associated with the connection and to return the results of monitoring to the network controller.

13. An arrangement according to Claim 11 and 12, wherein the network controller is arranged to provide a visual output of a representation of each network node and the transit delay data associated with that network node.

14. An arrangement according to any of Claims 11 to 13, wherein the network controller is arranged to determine a measure of transit delay between network nodes.

15. A switch according to any of Claims 1 to 9 including means for inputting a universal time reference, whereby transit times between similar switches can be correlated.

16. A method of obtaining a measure of a transit delay within a data switch comprising:

applying a time stamp to a cell arriving at the switch prior to switching by a switch core;

switching the cell within the switch core;

determining a measure of said delay by reading the time stamp after switching by the switch core.

17. A switch or a network substantially as any one herein described, or as illustrated in either of the accompanying drawings.

The Patent Office

14

INVESTOR IN PEOPLE

Application No: GB 9821554.4 **Examiner:** Richard Howe
Claims searched: 1-17 **Date of search:** 2 February 2000

Patents Act 1977
Search Report under Section 17

Databases searched:

UK Patent Office collections, including GB, EP, WO & US patent specifications, in:

UK Cl (Ed.R): H4K (KTK) ; H4P (PPS)

Int Cl (Ed.7): H04L (12/56)

Other: Online : wpi ; epodoc ; japio

Documents considered to be relevant:

Category	Identity of document and relevant passage		Relevant to claims
X	GB 2 261 799 A	(Dowty) - see abstract and whole document	1,16 at least
X	EP 0 544 454 A2	(Cray Communications) - see abstract and whole document	1,16 at least
X,&	EP 0 540 028 A2	(NEC Corporation) - see abstract and whole document	1,16 at least
X,&	US 5 485 457	(NEC Corporation) - see abstract and whole document	1,16 at least
X	US 5 337 308	(NEC Corporation) - see abstract and whole document	1,16 at least

X Document indicating lack of novelty or inventive step
Y Document indicating lack of inventive step if combined with one or more other documents of same category.
& Member of the same patent family
A Document indicating technological background and/or state of the art.
P Document published on or after the declared priority date but before the filing date of this invention.
E Patent document published on or after, but with priority date earlier than, the filing date of this application.